Conference Abstract

Biodiversity Data Curation: South Africa Goes Online

Willem Coetzer‡, Alexandra Holland§, Ian Engelbrecht|

‡ South African Institute for Aquatic Biodiversity, Grahamstown, South Africa
§ Albany Museum, Grahamstown, South Africa
| South African National Biodiversity Institute, Pretoria, South Africa

Abstract

The South African Institute for Aquatic Biodiversity (SAIAB) operates several research platforms, which may be used by the broader South African research community (e.g. a marine research vessel and a remotely operated underwater vehicle). SAIAB's Enterprise-grade data centre, along with expertise in systems administration and biodiversity information management, allow the institute to offer a Biodiversity Information Management Platform.

Data hosted by SAIAB is replicated across three data centres, with each centre being at least 250m apart and operating independently. Infrastructure at two data centres replicates in real time, forming a high availability cluster. The third datacentre is dedicated to storing backups. High-capacity tape backup will be added in the near future. As an additional measure, cloud storage is used to store daily extracts of Specify databases, which are retained for one year.

In the first instance, the Platform aims to provide SAIAB researchers and associates with biodiversity data curation services. This begins with support for the SAIAB Collections Division, to ensure that voucher specimens, tissue samples and associated media are accurately catalogued and can be easily retrieved. Biodiversity data curation is broader than this. It also means that any biodiversity data/metadata (records of species, events, occurrences/observations and traits) can potentially be curated using Specify Software,

and standardised and published (subject to relevant policies) to the GBIF Data Portal using the GBIF Integrated Publishing Toolkit. The use of Specify Software to curate biodiveristy data that do not represent voucher specimens (e.g. underwater images and video) is a new research project within SAIAB, which has the potential to be extended beyond SAIAB.

A new national initiative, the Natural Science Collections Facility (NSCF), was launched in 2017 to reinvigorate natural science museums across the country, to halt deterioration of specimens and improve capacity for specimen and data curation.

In support of the NSCF, the SAIAB platform is offered to natural science museums in South Africa (excluding herbaria, which are all part of or affiliated with SANBI, and therefore accommodated by a different system). Each museum will be provided with a webserver, Specify 7 database, Specify web portal and IPT server.

In offering this platform to the broader South African Biodiversity Science community, SAIAB is primarily motivated by the potential for collaborative research in capacity development for biodiversity data curation / information management, using Specify Software. The first research project will examine participating museums' capacity to use the Specify Workbench sustainably, to import new voucher/occurrence records generated by fieldwork. The requisite training to enhance this potential will be provided.

The Natural Science Collections Facility (NSCF) is an important collaborator in the context of enhancing the general state of South Africa's specimen collections, and the Specify Collections Consortium is an important collaborator, specifically for support.

Keywords

biodiversity data curation, capacity development, biodiversity information management

Presenting author

Willem Coetzer